Published in final edited form as:
Biometrika. 2013 September 1; 100(3): 709–726. doi:10.1093/biomet/ast011.

Robust analysis of semiparametric renewal process models

Feng-Chang Lin, **Young K. Truong**, and **Jason P. Fine**
Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A

Feng-Chang Lin: flin@bios.unc.edu; Young K. Truong: truong@bios.unc.edu; Jason P. Fine: jfine@bios.unc.edu

Summary

A rate model is proposed for a modulated renewal process comprising a single long sequence, where the covariate process may not capture the dependencies in the sequence as in standard intensity models. We consider partial likelihood-based inferences under a semiparametric multiplicative rate model, which has been widely studied in the context of independent and identical data. Under an intensity model, gap times in a single long sequence may be used naively in the partial likelihood with variance estimation utilizing the observed information matrix. Under a rate model, the gap times cannot be treated as independent and studying the partial likelihood is much more challenging. We employ a mixing condition in the application of limit theory for stationary sequences to obtain consistency and asymptotic normality. The estimator's variance is quite complicated owing to the unknown gap times dependence structure. We adapt block bootstrapping and cluster variance estimators to the partial likelihood. Simulation studies and an analysis of a semiparametric extension of a popular model for neural spike train data demonstrate the practical utility of the rate approach in comparison with the intensity approach.

Keywords

Block bootstrap; Mixing condition; Neurophysiology; Partial likelihood; Single sequence; Stationary limit theory

1. Introduction

In Cox's modulated renewal process model (Cox, 1972a), the conventional intensity function, $\lambda(t) = \lim_{\delta\to 0^+} \delta^{-1}\,\mathrm{pr}\{N(t+\delta) - N(t) = 1 \mid \mathcal{H}_t\}$, *is* assumed to satisfy

$$\lambda(t)=\lambda_0(V_t)\exp\{\beta^{\mathrm{T}} Z(t)\}. \quad (1)$$

Here $N(t)$ is a point process with jumps at the event times $\{T_i\}_{i\geq 1}$, and $\mathcal{H}_t$ is the history up to time t, including a length-q column vector $Z(t)$ of covariate processes; that is, $\mathcal{H}_t$ represents all information observed up to time t, which includes $N(t)$ and $Z(t)$ as well as, potentially, additional variables that do not enter model (1). In (1), β is a column vector of q parameters and λ_0 is an unspecified baseline intensity function. The backward recurrence time is defined by $V_t = t - T_{N(t-)}$, where $T_n = \inf\{t\colon N(t) \geq n\}$ and $T_{N(t-)}$ records the time of the newest event before t.

Supplementary material: Supplementary material available at *Biometrika* online includes proofs of consistency of the variance estimators, extra simulation results, and exploratory analysis of the spike train data.

For model (1) to be applicable, one has to condition on $\mathcal{H}_t$, including the whole history of events and the covariate process up to t. Of course, the dependence of the current event on previous events may not be adequately captured by $Z(t)$, and weaker modelling assumptions may be desirable. We propose a rate function, defined by $\lambda_{\mathrm{m}}(t) = \lim_{\delta\to 0^+} \delta^{-1}\, \mathrm{pr}\{N(t + \delta) - N(t) = 1 | V_t, Z(t)\}$. This can be modelled similarly to (1), with

$$\lambda_{\mathrm{m}}(t)=h_0(V_t)\exp\{\beta^{\mathrm{T}}Z(t)\}, \quad (2)$$

where the covariate effect β is proportional to the baseline rate function h_0. The rate function is the average event occurrence conditional only on current covariate values and the elapsed time from the newest event; we have (1) implies (2) when $\lambda_0 = h_0$, but not vice versa.

A similar definition of the rate function has been investigated for n independent and identical short sequences (Pepe & Cai, 1993; Lawless & Nadeau, 1995; Lin etal., 2000; Dabrowska & Ho, 2006). By short we mean that the number of jumps is assumed to be finite in each sequence, while the number of sequences n is assumed to grow. To our knowledge, such rate models have not been studied for data from a single long sequence, with the number of events, n, assumed to grow. It is worth noting that the effect of a time-independent covariate cannot be identified using data from a single long sequence, unlike with n independent and identical short sequences. Challenges in studying data from a single long sequence are well documented and distinct from those for short sequences; for an overview, see Chapter 7 in Daley & Vere-Jones (2003). The development of inferences requires careful consideration of the underlying correlation structure of the process.

Long single sequences occur widely in point process applications, for example, earthquake prediction in seismology, neural firing patterns in neurophysiology, and epidemic models in infectious disease monitoring. We consider data from an ensemble of neurons, in which the spiking probability of a target neuron is affected by concurrent peer cells. A point process framework proposed by Truccolo et al. (2005) is based on a parametric model for the conditional intensity that uniquely characterizes the distribution of a single spike train. Using the spiking history of peer neurons as covariates, the conditional intensity at a target neuron follows a multiplicative form, with parametric baseline intensity. We propose a natural generalization of this model via Cox's semiparametric formulation (Cox, 1972b), with inferences in the long single realization set-up following the general results of Lin & Fine (2009). Empirical findings such as those in § 5 suggest a highly nonlinear baseline, which may be difficult to model. Further complications arise, since the spiking activities of peer neurons may not fully explain the firing pattern of the target neuron. In reality, spike train models might best be viewed as descriptions of neuronal associations in a small set of neurons, rather than representations that fully capture the underlying spike train network, which may involve a large number of neurons. A marginal approach, such as the rate model, is ideal for data of this type, in that it only requires information on the current values of the selected covariates, for example the timing of previous spikes of the peer cells.

With λ_0 in (1) unspecified, Cox (1972b) proposed a partial likelihood for β, naively treating gap times from $N(t)$ as being independent. Oakes & Cui (1994) pointed out that there are two problems when using a partial likelihood for inference about β in (1). The first is that the partial likelihood reorders the time scale so that the counting processes $N_i(x) = I(X_i \le x)$, where $X_i = T_i - T_{i-1} (i = 1, \ldots, n)$, cannot be meaningfully defined with respect to a common filtration. The second is that $(N_i, Z_i)^{\mathrm{T}}$, where $Z_i(x) = Z(T_{i-1} + x)$ $(i \ge 1)$, may be unconditionally dependent as a result of correlations in $Z(t)$. Hence, martingale theory for n independent and identical sequences is not directly applicable, because $Z_j(x)$ is not predictable with respect to $\mathcal{H}_{T_{i-1}+x}$ when $j > i$. Oakes & Cui (1994) argued that the score function of the partial likelihood can be approximated using independent and identical comparison processes which have the same marginal distribution as $(N_i, Z_i)^{\mathrm{T}}$. Under this approximation, asymptotic properties of the estimator follow from conventional limit theory. Inferences can proceed as if the data were independent and identical, with variance estimation utilizing the observed information matrix.

Another theoretical approach originates from Pons & de Turckheim (1988), who studied Cox's model with periodic baseline intensity using a single realization under certain ergodicity conditions. Similar ideas could be applied to the modulated renewal model (1), with ergodicity conditions assumed such that unpredictable processes converge uniformly to deterministic limits which can be substituted in the estimating equation. After transforming to the original time scale, martingale theory is applicable to the approximate estimating function, yielding identical results to those of Oakes & Cui (1994). Lin & Fine (2009) adapted this approach to a general martingale estimating equation set-up for semiparametric intensity models.

There is a rich literature on nonparametric estimation of the interarrival distribution function from a sequence of dependent random variables, focused on either uncensored observations (Bagai & PrakasaRao, 1991; Yu, 1993) or possibly censored observations (Cai & Roussas, 1998; Cai, 1998, 2001; Leonenko & Sakhno, 2001). Kaplan–Meier-type estimation was proposed for dependent sequences, with kernel smoothing applied to obtain the corresponding density or hazard function. There have also been fruitful developments on an intensity-based dynamic model which includes event history as covariates to model the dependency between recurrent events (Aalen et al., 2004; Gandy & Jensen, 2004; Fosen et al., 2006; Borgan et al., 2007). Our rate function strategy can accommodate the event history by incorporating the information in the covariate. The rate-based dynamic model robustifies inferences relative to the intensity dynamic model, which could be important when the dynamic formulation is misspecified, as might occur if other unobserved peer cell histories are omitted from the model for the target neuron. This is illustrated in the simulations in § 4. Including the dynamic covariates generally improves the model fit, but doing so changes the interpretation of the coefficients for the nondynamic covariates. These practical issues are explored in the neuronal data analysis in § 5.

An estimating function similar to the partial likelihood score function is popular for proportional rate function analyses with independent and identically distributed observations (Pepe & Cai, 1993; Lin et al., 2000). This motivated us to study the same estimating function to estimate β in model (2) with a single long realization. It is unclear how to generalize the comparison process approach in Oakes & Cui (1994). Given the dependence among the gap times, it seems unlikely that the estimator is asymptotically equivalent to an estimator based on independent data. Moreover, under model (2), the event sequence is no longer defined by an intensity, so the ergodic martingale approach in Lin & Fine (2009) is not applicable. Similarly to Lin & Fine (2009), we posit certain conditions on the dependence structure of the process. However, as shown in § 3·1, the theoretical developments and resulting inferences are much more challenging than in Oakes & Cui (1994), Pons & de Turckheim (1988) and Lin & Fine (2009), owing to the lack of either an independence property or a martingale approximation for the estimating function.

2. Definitions and Model Assumptions

2·1. Definition of the estimators

We assume that the true value of β, denoted by β_0, falls into a bounded subset $\mathcal{B}$ of $\mathbb{R}^q$, and that the true baseline rate function of H, $H_0 = \int h_0$, belongs to a family of measures $\mathcal{L}$ on $\mathbb{R}_+$ having a Radon–Nikodym derivative with respect to Lebesgue measure. Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{P}$ be a family of probability measures on $(\Omega, \mathcal{F})$. Assume that a history $\mathcal{H}_t$ $(t \geq 0)$ and the smallest σ-algebra spanned by the current value of covariate $Z(t)$ are both P-complete sub-σ-algebras of $\mathcal{F}$ for any P in $\mathcal{P}$. We observe a single realization of $N(t)$ satisfying (2), along with information in the filtration. One should recognize that in a dynamic model, the covariate $Z(t)$ may contain information in $N(t)$.

Let T_i $(i \geq 1)$ be the associated event times of $N(t)$, and define $T_0 = 0$. Let the gap time $X_i = T_i - T_{i-1}$ and positive constant τ satisfy $\mathrm{pr}(X_i > \tau) > 0$ for each i. In theory, τ defines the upper bound on the support of the gap time beyond which information is not utilized. In practice, our numerical studies show that letting $\tau = \max\{X_1, \ldots, X_n\}$ performs well when including all information in $X_i (i = 1 \ldots, n)$. Let the counting process $N_i(x) = N(T_{i-1} + x) - N(T_{i-1})$ record the number of events in $(T_{i-1}, T_{i-1} + x]$, and define the covariate process as $Z_i(x) = Z(T_{i-1} + x)$.

Under model (2), let $\mathrm{d}\Lambda_{\mathrm{m},i}(\beta, x) = I(X_i \geq x)h_0(x)\exp\{\beta^{\mathrm{T}}Z_i(x)\}\mathrm{d}x$, and let $\mathrm{d}M_i(\beta, x) = \mathrm{d}N_i(x) - \mathrm{d}\Lambda_{\mathrm{m},i}(\beta, x)$. We propose to estimate β_0 with the root of the score function

$$U_{\mathrm{m}}(\beta)=\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(\upsilon)-\overline{Z}(\beta,\upsilon)\}\mathrm{d}N_i(\upsilon)=\sum_{i=1}^{n}\int_0^{\tau}\{Z_i(\upsilon)-\overline{Z}(\beta,\upsilon)\}\mathrm{d}M_i(\beta,\upsilon), \quad (3)$$

where $\overline{Z}=S_n^{(1)}/S_n^{(0)}$ with $S_n^{(k)}(\beta,x)=n^{-1}\sum_{j=1}^{n}I(X_j \geq x)Z_j(x)^{\otimes k}\exp\{\beta^T Z_j(x)\}$; here $a^{\otimes k}$ is equal to 1, a or aa^{T} for $k = 0$, 1 or 2, respectively. The function $U_{\mathrm{m}}(\beta)$ is identical to the score function of a naive partial likelihood based on the gap times and leads to an identical estimator, $\hat{\beta}_n$. In the following, we develop valid inferences for when model (2) holds but model (1) does not.

Estimation of H_0 may be useful for variance estimation for $\hat{\beta}_n$, as well as for assessing the shape of the baseline rate function. A Breslow-type estimator of H_0 is

$$\widehat{H}_0(\widehat{\beta}_n,x)=n^{-1}\sum_{i=1}^{n}\int_0^{x}S_n^{(0)}(\widehat{\beta}_n,\upsilon)^{-1}\mathrm{d}N_i(\upsilon). \quad (4)$$

For brevity, we write $\hat{H}_0(\hat{\beta}_n, x)$ as $\hat{H}_0(x)$.

2.2. Conditions for convergence

Although the estimating equations are identical under the rate and intensity model assumptions, the statistical inferences depend heavily on the true model. If model (1) holds, then the process $M_i(\beta_0, x)$ in (3) would be a martingale, since it correctly conditions on the whole filtration. However, if only model (2) holds, then $M_i(\beta_0, x)$ is no longer a martingale, since it only conditions on part of the history. Thus, the conventional limit theory in Andersen & Gill (1982) for martingale sequences is not applicable. Furthermore, the dependence between the processes M_i cannot generally be ignored in a long single sequence under model (2) as it can with independent sequences.

The large-sample properties of $\hat{\beta}_n$ must explicitly acknowledge the dependence structure of the sequence. As commonly employed with dependent processes, a mixing condition will facilitate the asymptotic analysis of the statistical procedures. We first define the strong mixing coefficient. For any two sub-σ-algebras $\mathcal{M}_1$, $\mathcal{M}_2 \subset \mathcal{F}$, a strong mixing coefficient of dependence is defined by

$$\alpha(\mathscr{M}_1, \mathscr{M}_2)=\sup|P_0(A\cap B)-P_0(A)P_0(B)|,$$

where $A \in \mathcal{M}_1$, $B \in \mathcal{M}_2$, and $P_0 \in \mathcal{P}$ is the probability associated with the true parameters β_0 and H_0. Given $n \in \mathbb{N}$, let $\alpha_n=\sup_{m\geq 1}\alpha(\mathscr{G}_1^m, \mathscr{G}_{m+n}^{\infty})$, where each $\mathscr{G}_k^l=\sigma(\xi_k, \ldots, \xi_l)(k \leq l)$ is a σ-algebra generated by possibly vector-valued random variables $\xi_k,\ldots, \xi_l$. A sequence ξ_i ($i \geq 1$) is strongly mixing if $\alpha_n \to 0$ as $n \to \infty$. In our context, ξ_i is a random variable whose value depends on X_i and $\{I(X_i \leq x)Z_i(x), 0 \leq x \leq \tau\}$, which is basically the random process Z marked between T_{i-1} and T_i ($i \geq 1$). Our main condition is stated as follows.

Condition 1—The marked random sequence $\{I(Xi \leq x)Z_i(x), 0 \leq x \leq \tau, X_i\}(i \geq 1)$ is strongly mixing and stationary.

Throughout the paper we will frequently refer to Conditions A1–A7 from the Appendix. Under Conditions 1 and A1–A7, we first establish that $S_n^{(k)}$ converges uniformly in x to a deterministic limit for each $\beta \in \bar{\mathcal{B}}$, a compact closure of $\mathcal{B}$; later we extend the result to uniform convergence in β. These results are critical to proving both consistency and asymptotic normality of $\hat{\beta}_n$. The pointwise result is stated in the following lemma.

Lemma 1—Suppose Condition 1 holds and that α_n satisfies

$$\sum_{n=1}^{\infty} n^{-1}(\log n)(\log\log n)^{1+\delta}\alpha_n<\infty$$

for some $\delta > 0$. Then

$$\sup_{x\in[0,\tau]}\left|S_n^{(k)}(\beta,x)-s^{(k)}(\beta,x)\right| \to 0$$

in probability for k = 0, 1,2 and every $\beta \in \bar{\mathcal{B}}$, where

$$s^{(k)}(\beta,x)=\lim_{n\to\infty}n^{-1}\sum_{j=1}^{n}E\{I(X_j \geq x)\}Z_j(x)^{\otimes k}\exp\{\beta^T Z_j(x)\}.$$

With β fixed the convergence follows from Corollary 2.1 in Cai & Roussas (1992) for a stationary strongly mixing sequence of random variables. The same argument can be applied to show that

$$\sup_{x\in[0,\tau]}\left|n^{-1}\sum_{i=1}^{n}N_i(x)-\phi(x)\right| \to 0$$

and

$$\sup_{x\in[0,\tau]}\left|n^{-1}\sum_{i=1}^{n}\Lambda_{\mathrm{m},i}(\beta_0,x)-\phi(x)\right|\to 0$$

in probability where $\phi(x)=\int_0^x s^{(0)}(\beta_0,\upsilon)\mathrm{d}H_0(\upsilon)$. Similarly, $n^{-1}\sum_{i=1}^{n}\int_0^x Z_i(\upsilon)\mathrm{d}N_i(\upsilon)$ converges uniformly to a continuous function $\int_0^x s^{(1)}(\beta_0,\upsilon)\mathrm{d}H_0(\upsilon)$.

Lemma 1 can be strengthened to obtain the uniform convergence of $S_n^{(k)}$ in β, using arguments similar to those in the proof of Lemma 2.3 in Pons & de Turckheim (1988).

Lemma 2—If Conditions 1 and A1 hold, then for $k = 0, 1,2$,

$$\sup_{|\beta-\beta_0|\le\rho}\sup_{x\in[0,\tau]}\left|S_n^{(k)}(\beta,x)-S_n^{(k)}(\beta_0,x)\right|\to 0$$

in probability when $\rho = o(n)$.

3. Asymptotic Properties

3. 1. Consistency and asymptotic normality of $\hat{\beta}_n$

The following theorem gives the consistency of $\hat{\beta}_n$.

Theorem 1—If Conditions 1, A2 and A3 hold, then $\hat{\beta}_n$ is consistent for β_0.

The proof requires only the pointwise convergence for each β in Lemma 1, and not uniform convergence with respect to β. This is the case because the partial likelihood used to construct $U_{\mathrm{m}}(\beta)$ is concave and has a pointwise deterministic limit with a unique maximizer at β_0.

In order to establish the limit distribution of $n^{1/2}(\hat{\beta}_n - \beta_0)$, we will employ a first-order Taylor series approximation. The key step is deriving the large-sample distribution of $n^{-1/2}U_{\mathrm{m}}(\beta_0)$. Under model (1), when the martingale central limit theorem is applicable, the limiting variance is equal to $\Omega(\beta_0)$, which is the limiting value of the average unit information per observation, as shown in Oakes & Cui (1994). Under model (2), when $(M_1, \ldots, M_n)^{\mathrm{T}}$ in $U_{\mathrm{m}}(\beta_0)$ are weakly dependent and there is no martingale structure, we can show that under certain regularity conditions, $n^{-1/2}U_{\mathrm{m}}(\beta_0)$ is asymptotically normal with mean zero and variance

$$\Phi(\beta_0)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}E\{\tilde{U}_i(\beta_0)\tilde{U}_j^{\mathrm{T}}(\beta_0)\},$$

where $\tilde{U}_i(\beta)=\int_0^\tau\{Z_i(\upsilon)-\overline{z}(\beta,\upsilon)\}\mathrm{d}M_i(\beta,\upsilon)$ and $\bar{z} = s^{(1)}/s^{(1)}$. This leads to the following theorem.

Theorem 2—If Conditions 1 and A1–A7 hold, then $n^{1/2}(\hat{\beta}_n - \beta_0)$ converges in distribution to a zero-mean normal variable with covariance matrix $\Sigma(\beta_0) = \Omega(\beta_0)^{-1}\Phi(\beta_0)\Omega(\beta_0)^{-1}$.

Under model (1), $\Phi(\beta_0) = \Omega(\beta_0)$ so that $\Sigma(\beta_0) = \Omega(\beta_0)^{-1}$, which is the same limiting variance of $\hat{\beta}_n$ as in Oakes & Cui (1994). When model (2) holds but model (1) does not, $\Phi(\beta_0) \neq \Omega(\beta_0)$ so that $\Sigma(\beta_0) \neq \Omega(\beta_0)^{-1}$. The simulation study in §4 shows that ignoring the dependence in variance estimation may result in estimators with poor performance. Valid inferences require estimating $\Phi(\beta_0)$ in the sandwich formula, which may be challenging due to complex dependencies in the gap time sequence.

3.2. Asymptotic distribution of $n^{1/2}\{\hat{H}_0(x) - H_0(x)\}$

We next show that the estimator (4) is uniformly consistent for $x \in [0, \tau]$ and converges weakly to a tight Gaussian process, as stated in Theorem 3 below.

Theorem 3—Under Conditions 1 and A1–A7, $\hat{H}_0(x)$ converges to $H_0(x)$ uniformly for $x \in [0, \tau]$. Furthermore, $n^{1/2}\{\hat{H}_0(x) - H_0(x)\}$ converges weakly to a zero-mean Gaussian process with continuous sample paths and covariance function

$$\gamma(x,y)=\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}E\{E\{\Xi_i(x)\Xi_j(y)\},$$

where $\Xi_i(x)=\int_0^x s^{(0)}(\beta_0,v)^{-1}\mathrm{d}M_i(\beta_0,v)-\Omega_0(\beta_0)^{-1}\tilde{U}_i(\beta_0)\int_0^x\overline{z}(\beta_0,v)\mathrm{d}H_0(v)$.

3.3. Variance estimation

The estimation for $\Sigma(\beta_0)$ relies on the estimation of $\Phi(\beta_0)$, since we can consistently estimate $\Omega(\beta_0)$ by $\hat{\Omega}(\hat{\beta}_n)$ where

$$\widehat{\Omega}(\beta)=\int_0^\tau\left\{S_n^{(2)}(\beta,v)-\frac{S_n^{(1)}(\beta,v)^{\otimes 2}}{S_n^{(0)}(\beta,v)}\right\}\mathrm{d}\widehat{H}_0(\beta,v).$$

If model (1) is applicable, then the dependence among the terms in U_{m} can be ignored and $\Phi(\beta_0)$ can be consistently estimated by $\widehat{\Phi}(\widehat{\beta}_n)=n^{-1}\sum_{i=1}^{n}\int_0^\tau\{Z_i(v)-\overline{Z}(\widehat{\beta}_n,v)\}^{\otimes 2}\mathrm{d}N_i(v)$, which is equal to $\hat{\Omega}(\hat{\beta}_n)$ as $n \to \infty$. Thus, $\Sigma(\beta_0)$ can be estimated by

$$\widehat{\sum}_0=\widehat{\Omega}(\widehat{\beta}_n)^{-1}. \quad (5)$$

When the rate model (2) holds but the intensity model (1) does not, sandwich variance estimation is needed. However, consistent estimation of $\Phi(\beta_0)$ is challenging because the dependence structure of $(M_1, \ldots, M_n)^{\mathrm{T}}$ may not be clear under a rate model. A common approach to inference with dependent processes is the block bootstrap (Lahiri, 2003). We now discuss adaptations of this approach to the partial likelihood-based analysis of model (2).

There are two possible choices of resampling units. The first is based on $U_i^N(\beta)=\int_0^\tau\{Z_i(v)-\overline{Z}(\beta,v)\}\mathrm{d}N_i(v)$, which can be calculated without estimating H_0. The second is $U_i^M(\beta)=\int_0^\tau\{Z_i(v)-\overline{Z}(\beta,v)\}\mathrm{d}\widehat{M}_i(\beta,v)$ where

$$\widehat{M}_i(\beta,x)=N_i(x)-\int_0^x I(X_i\geq \upsilon)\exp\{\beta^T Z_i(\upsilon)\}\mathrm{d}\hat{H}_0(\beta,\upsilon),$$

which employs the estimator $\hat{H}_0(x)$ from § 3.2. Both resampling pools are of interest, since $n^{-1/2}U_{\mathrm{m}}(\beta_0)$ is asymptotically equivalent to $n^{-1/2}\Sigma_i \tilde{U}_i(\beta_0)$. The latter quantity has the same limit distribution as $n^{-1/2}\sum_i U_i^N(\beta_0)=n^{-1/2}\sum_i U_i^M(\beta_0)$, with covariance matrix $\Phi(\beta_0)$. Essentially, U_i^M and U_i^N are surrogates for the unknown $\tilde{U}_i$. In practice, one replaces β_0 with $\hat{\beta}_n$ in U_i^N and U_i^M and bootstraps these score residuals, denoted generically by $U_i(\hat{\beta}_n)$ $(i \geq 1)$. The question is how to resample these residuals to obtain valid variance estimates.

For sequences of strongly mixing random variables, the block bootstrap of Künsch (1989) can be useful (Politis & Romano, 1992; Shao & Yu, 1993; Peligrad, 1998). Under Condition A1, each $\tilde{U}_i(\beta_0)$ $(i \geq 1)$ is a strongly mixing sequence which is asymptotically equivalent to $U_i(\beta_0)$. Assuming $E\{U_i(\beta_0)\}^{\otimes 2}<\infty$, an extension of the circular block bootstrap variance estimator applied to $U_i(\hat{\beta}_n)$ can be proved to be consistent for $\Phi(\beta_0)$. Let k and l be two integers such that $n = kl$. By Peligrad (1998, Theorem 2.1), when the length of block l satisfies $l^2/n \to 0$ as $n \to \infty$, the blockwise bootstrapped estimator of the variance of $n^{-1/2}U_{\mathrm{m}}(\beta_0)$ is $\widehat{\Phi}_{\mathrm{b}}(\widehat{\beta}_n)=k^{-1}\sum_{i=1}^{n}\overline{U}_{li}(\widehat{\beta}_n)^{\otimes 2}$, where $\overline{U}_{li}(\widehat{\beta}_n)=l^{-1}\sum_{j=i}^{i+l-1}U_j(\widehat{\beta}_n)$ from an augmented sequence

$$\{U_1(\widehat{\beta}_n),\ldots,U_n(\widehat{\beta}_n),U_{n+1}(\widehat{\beta}_n),\ldots,U_{n+l-1}(\widehat{\beta}_n)\}$$

with $U_j = U_{j-n}$ when $j > n$. The variance estimators are

$$\widehat{\sum}_{\mathrm{b}}^{N}=\widehat{\Omega}(\widehat{\beta}_n)^{-1}\widehat{\Phi}_{\mathrm{b}}^{N}(\widehat{\beta}_n)\widehat{\Omega}(\widehat{\beta}_n)^{-1} \quad (6)$$

and

$$\widehat{\sum}_{\mathrm{b}}^{M}=\widehat{\Omega}(\widehat{\beta}_n)^{-1}\widehat{\Phi}_{\mathrm{b}}^{M}(\widehat{\beta}_n)\widehat{\Omega}(\widehat{\beta}_n)^{-1} \quad (7)$$

for resampling units U_i^N and U_i^M, respectively.

There is an extensive literature on bootstrapping estimating functions evaluated at the estimated parameter, when the data are independent and identically distributed (Hu & Zidek, 1995; Hu & Kalbfleisch, 2000), independent but not identically distributed (Lele, 1991a, 1991b, 2003), or dependent random variables for method of moments estimators (Hall & Horowitz, 1996). The main condition is that the plug-in estimators should consistently estimate the parameters of interest. Following earlier work, one can show that $U_i(\hat{\beta}_n)$ will yield correct inferences, asymptotically, since the parameter estimators are consistent under the strong mixing condition.

Under more restrictive assumptions on the dependence structure, it may be possible to develop simple, plug-in variance estimators. Here, we consider lagged dependence across clusters, similar to the set-up in Andrews (1991) and Hansen (1992).

Let T_{ij} ($1 \le i \le K;\ 1 \le j \le K_i$) be the occurrence time of the jth event in the ith cluster; define $T_{10} = 0$ and $T_{i0} = T_{(i-1)K_{i-1}}$ for $i \ge 2$. The total number of events is $n = \sum_{i=1}^{K} K_i$. Define $X_{ij} = T_{ij} - T_{i(j-1)}$ to be the sequence of gap times for the jth event in the ith cluster, and let Z_{ij} be the corresponding covariate processes. We assume that gap times in cluster i have a common correlation structure and that cluster i may be correlated with earlier clusters $i - 1, \ldots, i - d + 1$ and later clusters $i + 1, \ldots, i + d - 1$, where d is a positive integer. There is no correlation for lags greater than $d - 1$. Further regularity conditions are discussed in Andrews (1991) and Hansen (1992) for certain regression models.

Under the clustered set-up, $U_{\mathrm{m}}(\beta)$ in (3) becomes

$$U_{\mathrm{m}}^{c}(\beta) = \sum_{i=1}^{K}\sum_{j=1}^{K_i}\int_0^{\tau}\{Z_{ij}(\upsilon) - \overline{Z}(\beta,\upsilon)\}\mathrm{d}N_{ij}(\upsilon),$$

where $N_{ij}(x) = N(T_{i(j-1)} + x) - N(T_{i(j-1)})$ and $\overline{Z}(\beta,x) = S_{\mathrm{c}}^{(1)}(\beta,x)/S_{\mathrm{c}}^{(0)}(\beta,x)$ with

$$S_{\mathrm{c}}^{(k)}(\beta,x) = K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{K_i} I(X_{ij} \ge x) Z_{ij}(x)^{\otimes k}\exp\{\beta^T Z_{ij}(x)\}\ (k = 0,1,2).$$

The solution $\hat{\beta}_n$ for $U_{\mathrm{m}}^{\mathrm{c}}(\beta) = 0$ here is the same as that from (3), but the variance estimator can be calculated without resampling.

Let

$$\begin{aligned}
\widehat{\Omega}_{\mathrm{c}}(\beta) &= \int_0^{\tau}\left\{S_{\mathrm{c}}^{(2)}(\beta,\upsilon) - \frac{S_{\mathrm{c}}^{(1)}(\beta,\upsilon)^{\otimes 2}}{S_{\mathrm{c}}^{(0)}(\beta,\upsilon)}\right\}\mathrm{d}\widehat{H}_{\mathrm{c}}(\beta,\upsilon),\\
\widehat{H}_{\mathrm{c}}(\beta,\upsilon) &= K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{K_i}\int_0^{\tau} S_{\mathrm{c}}^{(0)}(\beta,\upsilon)^{-1}\mathrm{d}N_{ij}(\upsilon),\\
\widehat{\Psi}_i(\beta) &= \sum_{j=1}^{K_i}\int_0^{\tau}\{Z_{ij}(\upsilon) - \overline{Z}(\beta,\upsilon)\}\mathrm{d}\widehat{M}_{ij}(\beta,\upsilon),\\
\widehat{M}_{ij}(\beta,\upsilon) &= N_{ij}(\upsilon) - \int_0^{\upsilon} I(X_{ij} \ge u)\exp\{\beta^T Z_{ij}(u)\}\mathrm{d}\widehat{H}_{\mathrm{c}}(\beta,u).
\end{aligned}$$

The estimator for $\Omega(\beta_0)$ is $\hat{\Omega}_{\mathrm{c}}(\hat{\beta}_n)$, and the theoretical quantity $\Phi(\beta_0)$ can be consistently estimated by $\widehat{\Phi}_{\mathrm{c}}(\widehat{\beta}_n) = K^{-1}\sum_{i=1}^{K}\widehat{\Psi}_i(\widehat{\beta}_n)^{\otimes 2}$ if the K clusters are mutually independent, so that $\Sigma(\beta_0)$ may be estimated by $\hat{\Sigma}_{\mathrm{c}} = \hat{\Omega}_{\mathrm{c}}(\hat{\beta}_n)^{-1}\,\hat{\Phi}_{\mathrm{c}}(\hat{\beta}_n)\hat{\Omega}_{\mathrm{c}}(\hat{\beta}_n)^{-1}$. If the dependence occurs only within $d - 1$ lags of clusters, the estimator can be computed with $\widehat{\Phi}_{\mathrm{c}}^{d}(\widehat{\beta}_n) = \sum_{j=-d+1}^{d-1}\widehat{\Gamma}_j(\widehat{\beta}_n)$, where $\widehat{\Gamma}_j(\beta) = K^{-1}\sum_{k=1}^{K-j}\widehat{\Psi}_k(\beta)\widehat{\Psi}_{k+j}^{\mathrm{T}}(\beta)$ if $j \ge 0$ and $\hat{\Gamma}_j = \hat{\Gamma}_{-j}$ if $j < 0$, so that $\Sigma(\beta_0)$ may be estimated by

$$\widehat{\sum}_{\mathrm{c}}^{d} = \widehat{\Omega}_{\mathrm{c}}(\widehat{\beta}_n)^{-1}\widehat{\Phi}_{\mathrm{c}}^{d}(\widehat{\beta}_n)\widehat{\Omega}_{\mathrm{c}}(\widehat{\beta}_n)^{-1}. \quad (8)$$

When we do not have clear information on whether an event belongs to a certain cluster or on the number of lags, $d - 1$, the estimating function (3) with block bootstrap variance estimation will provide valid inferences while the plug-in estimator requiring clustering assumptions is not computable. On the other hand when we have clear information on the structure of the long single sequence, the lag-($d - 1$) variance estimator described above is computationally simpler and may perform better for small samples, as illustrated empirically in §4. Proofs of the consistency of both variance estimators can be found in the Supplementary Material.

4. Simulation Studies

We generated the renewal process using a clustered data set-up as follows. Let $N_{r,\delta}(t) = N\{T_{C(t)0} - (r - 1)\delta\} - N(T_{C(t)0} - r\delta)$ count the number of events during the time interval $(T_{C(t)0} - r\delta, T_{C(t)0} - (r - 1)\delta]$, where r is a positive integer, $\delta > 0$, and $C(t) = \sup\{i: T_{i0} \le t\}$ is the number of clusters before t. The number of events $N(t)$ is identically zero for $t \le 0$. Our simulated data were generated by an intensity model

$$\lambda(t)=W(t)\lambda_0(V_t)\exp\left\{\gamma_0 Z(t)+=\sum_{r=1}^{R}\gamma_r N_{r,\delta}(t)\right\}, \quad (9)$$

where $W(t) = W_i I(T_{i0} \le t < T_{(i+1)0})$ is an unobserved cluster-level frailty process which is independent of the observed covariate process $Z(t)$, and R is the number of intervals lagging back from $T_{C(t)0}$ in the process history.

Conditionally on $W(t)$, $Z(t)$ and the history of events, a multiplicative intensity model holds. However, since $W(t)$ is unobserved the model being fitted to the observed data is unconditional on $W(t)$. It is known that a random-effects proportional intensity model does not generally satisfy the multiplicative model unconditionally on the random effects (Hougaard, 2000). One can show that if $\gamma_r = 0$ in (9) and the marginal distribution of W_i is positive stable with parameter α_0, or if $\gamma_r \ne 0$ and the W_i are independently drawn from a positive stable distribution with parameter α_0, then a rate model holds with

$$\lambda_{\mathrm{m}}(t)=h_0(V_t)\exp\left\{\beta_0 Z(t)+=\sum_{r=1}^{R}\beta_r N_{r,\delta}(t)\right\} \quad (10)$$

where $h_0(V_t)=\alpha_0\lambda_0(V_t)\Lambda_0(V_t)^{\alpha_0-1}$, $\Lambda_0(V_t)=\int_0^{V_t}\lambda_0(\upsilon)\mathrm{d}\upsilon$ and $\beta_r = \alpha_0\gamma_r$.

In the first scenario, the results for which are shown in Table 1, we assumed that $\gamma_r = 0$ for $r \ge 1$ and that $W_i=\sum_{j=i}^{i+d-1}Y_j$, where the Y_j are drawn from a positive stable distribution with Laplace transformation $E\{\exp(-sY_j)\}=\exp(-d^{-1}s^{\alpha_0})$. Specifically, if $d > 0$, the frailty variable W_i marginally follows a positive stable distribution (Hougaard, 2000) and has correlations that satisfy the lag-($d - 1$) assumption of Andrews (1991) and Hansen (1992). In the second scenario, the results for which are reported in Table 2, we let $\gamma_r = 10^{-r}$ for $r \ge 1$, $\delta = 0{\cdot}2$ and $R = 10$, and draw the Wi independently from a positive stable distribution with Laplace transformation $E\{\exp(-sY_j)\} = \exp(-s^{\alpha_0})$, which implies not only a dynamic model with event history as covariates but also a shared frailty model where gap times in the same cluster are dependent. For both scenarios, we assume $Z(t) = Z_{ij}I(T_{i(j-1)} \le t < T_{ij})$, with the Z_{ij} independently drawn from a uniform distribution between 0 and 1. We initiated a new cluster with fixed probability 0-25 in the first scenario, but with different probabilities in the second scenario in order to investigate the impact of the average cluster size. The baseline intensity was taken to be $\lambda_0(V_t) = 1$ for all t. The regression coefficient of $Z(t)$ in model (9) is $\gamma_0 = -\log(5)$. The parameter for the positive stable distribution is α_0 = 0·75 or 0·5, where a smaller value of α_0 represents greater dependence between gap times. The lags in Table 1 are d = 0, 2 and 5, with sample sizes of n = 200, 400, 1000 and 2000, where the results for d = 0 correspond to simulations under an intensity model with $W_i = 1$ for all i. The sample sizes in Table 2 are n = 400, 1000, 2000 and 4000, with average cluster sizes of 5 and 10. We report bias, empirical variance, coverage probability, and various variance estimators, as given in (5)–(8). We also simulated the data in the case where Z_i follows a first-order autoregressive model. The results are similar and are presented in the Supplementary Material.

As Table 1 shows, the variance estimator $\hat{\Sigma}_0$ defined in (5), where gap times are naively treated as independent variables, is only valid under an intensity model. With d = 0, the point estimation is unbiased the variance estimation is close to the empirical variance, and the coverage probability for a nominal 0·95 Wald-type confidence interval is close to 0.95. As d increases, the variance estimator may substantially underestimate the true variance, since the covariance from unobserved W is ignored and the coverage probability may be low. The sandwich variance estimator using lag-(d − 1) cluster assumptions, namely $\widehat{\sum}_{\mathrm{c}}^{d}$ defined in (8), is more robust. For $d \geq 2$, $\widehat{\sum}_{\mathrm{c}}^{d}$ properly accounts for dependence across clusters, exhibits some underestimation for small n, but performs well as n increases, achieving close to the 0·95 nominal level. For the block bootstrapped estimators, we used a block length of $l = n^{2/5}$, which satisfies the condition $l^2/n \to 0$. Upon exploring other choices we found that variance estimation is stable for l between $n^{2/5}$ and $n^{1/2}$. In general, the estimator $\widehat{\sum}_{\mathrm{b}}^{N}$ defined in (6) agrees more closely with the empirical variance than does $\widehat{\sum}_{\mathrm{b}}^{M}$ defined in (7). There is some evidence of underestimation of the variance and subnominal coverage probabilities for small n, with improved performance as n increases. The block bootstrap estimator with the best overall performance performs almost the same as $\widehat{\sum}_{\mathrm{c}}^{d}$.

In Table 2, the naive variance estimator $\hat{\Sigma}_0$ has an acceptable coverage probability when the cluster size is relatively small, although underestimation is evident due to unmodelled correlations within clusters. With a larger cluster size and higher dependence between gap times, the naive variance estimator may have undesirable coverage. Robust variance estimators such as $\widehat{\sum}_{\mathrm{c}}^{d}$ and $\widehat{\sum}_{\mathrm{b}}^{N}$ are close to the empirical variance and achieve the nominal coverage levels as the sample size increases, similarly to the first scenario.

5. Neural Spike Train Data Analysis

Let (0, T] denote the observation interval, which may be partitioned into small subintervals $(t_{k-1}, t_k]$ (k= 1, …, K), each of length $\delta = TK^{-1}$. The discrete-time representation of the conditional intensity of a target spiking process can be expressed as

$$\lambda(t_k)=\exp\left(\beta_0+\sum_{c=1}^{C}\sum_{r=1}^{R}\beta_r^c\Delta N_{k-r}^c\right), \quad (11)$$

where the first summation is over C peer cells in the ensemble and the second summation is over their spiking history up to the Rth lag from t_{k-1} (Truccolo et al., 2005). More precisely, $\Delta N^c_{k-r} \equiv N^c(t_{k-r}) - N^c(t_{k-r-1})$, where $N^c(\cdot)$ refers to peer cell spiking counts and $\beta = (\beta_0, \ldots, \beta^C_R)^{\mathrm{T}}$ is a length-$(C \times R + 1)$ column vector of regression parameters. Ideally, the subinterval length, δ, is small enough so that ΔN^c_{k-r} is a binary indicator. Thus the coefficient β^c_r reflects the lagged synchrony between the target cell and the peer cell. However, owing to the rapid growth of the number of parameters in model (11) with longer peer history information, an alternative conditional intensity model with lower time precision may also be practicable; this is defined by

$$\lambda(t_k) = \exp\left(\beta_0 + \sum_{c=1}^{C}\sum_{r=1}^{R} \beta^c_r \Delta N^c_{k,r,W}\right), \quad (12)$$

where $\Delta N^c_{k,r,W} \equiv N^c(t_{k-1-(r-1)W}) - N^c(t_{k-1-rW})$ counts the number of spikes in the peer cell c sequence during the time interval $(t_{k-1-rW}, t_{k-1-(r-1)W}]$. In model (12), the length W is such that $W \gg \delta$, thereby limiting the parameter dimension.

Clearly, both of the models (11) and (12) have multiplicative intensity with constant baseline $\exp(\beta_0)$. As discussed in § 1, we propose to generalize such models by incorporating an unspecified time-varying baseline and weakening the intensity assumption via the rate model. Accordingly, the continuous-time spike train rate model is

$$\lambda_{\mathrm{m}}(t) = h_0(V_t)\exp\left\{\sum_{c=1}^{C}\sum_{r=1}^{R} \beta^c_r Z^c_{r,W}(t)\right\}, \quad (13)$$

where $Z^c_{r,W}(t) = N^c\{t - (r-1)W\} - N^c(t - rW)$ is defined similarly to $\Delta N^c_{k,r,W}$ but on a continuous time scale. Including the target cell spiking history as a covariate leads to a dynamic model, which could possibly enhance goodness-of-fit relative to a model that does not include the target neuron's spiking history. Here we assume that the effect coming from the spiking history of the target cell is also proportional to the baseline rate function. The augmented rate model is

$$\lambda_{\mathrm{m}}(t) = h_0(V_t)\exp\left\{\sum_{c=1}^{C}\sum_{r=1}^{R} \beta^c_r Z^c_{r,W}(t) + \sum_{r=1}^{R} \beta^s_r Z^s_{r,W}(t)\right\}, \quad (14)$$

where the superscript indicates the target cell. Both models, with or without target cell firing history, may be of scientific interest and provide complementary information, even though the model with target cell history, (14), will generally yield better prediction than the restricted model (13). The model without target cell history provides unadjusted estimates of peer cell effects that can be compared to adjusted effects from the model with target history, which may potentially obscure the peer cells (Truccolo et al., 2005).

In studying how motivational salience is processed in the brain, Lin & Nicolelis (2008) tested whether the salience is encoded by ensemble bursting of non-cholinergic basal forebrain single neurons in behaving rats by simultaneously recording the activity of many basal forebrain neurons with movable multi-electrode bundles. In particular, they investigated whether motivationally salient cues predicting reward or punishment in a go/no-go task, as well as the reward and punishment themselves, elicited bursting responses in basal forebrain neurons. One objective is to identify those neurons which initiate the firing activity. To do so, an ensemble of three neurons, s1, s2 and s3, will be considered. We treated s1 as the target neuron and modelled its spiking activity using spike counts from the other cells as covariates. The number of spikes before the end of the observation time, T = 11271 seconds, for neurons s1, s2 and s3 were 226457, 65 847 and 162 146, respectively.

Some exploratory analysis and model selection issues related to the interval width are described in the Supplementary Material. Figure 1 shows the point estimates from our partial likelihood approach using 0-95 Wald-type confidence intervals with $R = 10$, with or without target spiking history as covariates. Naive and robust variance estimates are provided where the former was computed by inverting an information matrix and the latter was computed using the block bootstrapping technique with block length $l = 60 \simeq n^{1/3}$. The two variance estimators are very different, with the standard errors derived under the rate model assumption generally being twice as large as those from naive estimates based on an intensity model assumption. The statistical significance of particular parameters may depend critically on the method of variance estimation, for example at the fourth- and sixth-second lags in peer cell 2 without target history, as shown in Fig. 1(c). Interestingly, as Fig. 1(a) and (c) show, without including the target history as covariates, peer cell 1 did not have a significant impact in the third second and peer cell 2 did not have a significant impact on the firing rate of the target neuron until the second second. At significant lags, both peer cells inhibited the firing rate of the target neuron and as expected the impact is stronger at more recent lags.

The impact of the target neuron's spiking history is shown in Fig. 2(a), where we have omitted the estimate of the coefficient for the first lag due to its magnitude, with $\widehat{\beta}_1^s = 0.0616$. In general, the target cell firing history has a positive effect on the future firing rate up to the fourth-second lag. The stimulatory effects diminish at longer lags, becoming statistically insignificant. On the other hand the effects of peer cell 1 appear somewhat dampened relative to the unadjusted model, with the second-second lag not statistically significant, as shown in Fig. 1(c). Interestingly, for peer cell 2 in Fig. 1(d), the first lag is statistically significant after the adjustment and comparable in magnitude to the second lag, which differs qualitatively from the unadjusted analysis. The goodness-of-fit of model (14) is shown in Fig. 2(b); this is compared with the goodness-of-fit of model (13) in the Supplementary Material. While both models fit the observed number of spikes reasonably well, model (14), with the target neuron spiking history as covariates, gives a significantly better fit than does model (13).

The estimation of the cumulative baseline rate function, $\hat{H}_0(x)$, defined in (4), is shown in Fig. 2(c) with 0.95 pointwise confidence intervals. A constant baseline rate model such as (11) or (12) may not be valid in this case, as the estimate departs from a straight line. Estimation of the baseline rate function is shown in Fig. 2(d), obtained by using locally weighted polynomial regression (Cleveland & Devlin, 1988) to smooth $\hat{H}_0(x)$. Interestingly, the baseline firing rate is extremely nonconstant, in violation of the usual neural spike train modelling assumptions (11) and (12). The firing rate is rather low just after firing, as in a refractory period, then increases steadily, reaching a maximum at about 60 milliseconds, after which the rate decreases gradually, to approach an asymptote after about 250 milliseconds.

6. Discussion

Violations of the mixing conditions may invalidate the variance estimators. Usually such issues arise in analyses of point process data, where some assumptions on the correlation structure are needed for the development of inferences. The mixing sequence approach has been widely used in application areas for dependent data (Bradley, 2005), and it would be of interest to develop empirical checks of the assumptions. As a start, one might investigate $k^{-1}\sum_{i=1}^{k} I(X_i \geq x) Z_i(x) (k=1, \ldots, n)$. If Lemma 1 holds, then these averages should converge as k increases. Similarly, the empirical correlations of the U_i may be employed to informally assess the strong mixing property required for bootstrapping. Neither of the above checks indicated strong evidence for violations in the case of the neural spike train data. Formal goodness-of-fit tests would be of practical use in evaluating these assumptions further, but such tests are beyond the scope of the present paper.

A main contribution of this paper is the development of methods for valid variance estimation under dependence assumptions that are weaker than those required for validity of the intensity model. Several approaches applicable to independent and identical sequences of dependent data might prove useful in this regard. Huang & Chen (2003) and Luo & Huang (2011) considered within-cluster dependence across independent clusters where the cluster structure is known a priori. Lin et al. (1993) utilized a conditional multiplier bootstrap which is generalizable to independent sampling units. Further work is needed to extend these approaches to dependent data from a single long sequence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Shih-Chieh Lin and Miguel Nicolelis for allowing us to use the neural spike train data. We also thank the editor and three referees for their helpful comments. F.-C. Lin and J. P. Fine were supported by the U.S. National Institutes of Health. Y. K. Truong was supported by the U.S. National Science Foundation.

Appendix

We assume the following regularity conditions to hold throughout the proofs:

Condition A1. $\sup_{\beta \in \bar{\mathcal{B}}} \sup_{x \in [0, \tau]} |I(X_1 \geq x) Z_1(x)^{\otimes k} \exp\{\beta^{\mathrm{T}} Z_1(x)\}|$ is integrable;

Condition A2. $s^{(0)}(\beta, x)$ is bounded away from zero on $\bar{\mathcal{B}} \times [0, \tau]$, and there exist continuous functions $s^{(k)}(\beta_0, x)$ on $[0, \tau]$ for $k = 0, 1, 2$;

Condition A3. $\Omega(\beta_0) = \int_0^{\tau} \{s^{(2)}(\beta_0, \upsilon)/s^{(0)}(\beta_0, \upsilon) - \bar{z}(\beta_0, \upsilon)^{\otimes 2}\} s^{(0)}(\beta_0, \upsilon) \mathrm{d}H_0(\upsilon)$ is positive definite;

Condition A4. $n^{1/2}\{ES_n^{(k)}(\beta_0, x) - s^{(k)}(\beta_0, x)\}$ is finite for $k = 0, 1, 2$ and for each x in a dense subset of $[0, \tau]$ including 0 and τ;

Condition A5. for $k = 0, 1,2$ and $n \geq 1$, $n^{1/2}\{S_n^{(k)}(\beta_0, x) - s^{(k)}(\beta_0, x)\}$ is tight on $[0, \tau]$;

Condition A6. $\sum_{n=1}^{\infty} \int_0^{\alpha_n} Q^2_{|\tilde{U}_1|}(\upsilon) \mathrm{d}\upsilon < \infty$, where $Q_W(\upsilon) = \inf\{x \geq 0: \mathrm{pr}(W > x) \leq \upsilon\}$ for any non-negative random variable W;

Condition A7. for $n \geq 1$, the sequence of processes $n^{-1/2}\sum_{i=1}^{n} M_i(\beta_0, x)$ is tight on $[0, \tau]$.

Proof of Theorem 1

Consider

$$\ell_{\mathrm{m}}(\beta)=\sum_{i=1}^{n}\int_0^\tau \beta^{\mathrm{T}} Z_i(\upsilon)\mathrm{d}N_i(\upsilon)-\sum_{i=1}^{n}\int_0^\tau \log\{nS^{(0)}(\beta,\upsilon)\}\mathrm{d}N_i(\upsilon)$$

and $U_{\mathrm{m}}(\beta) = \partial\ell_{\mathrm{m}}(\beta)/\partial\beta$. Let $R(\beta) = n^{-1}\{\ell_{\mathrm{m}}(\beta) - \ell_{\mathrm{m}}(\beta_0)\}$, which equals

$$n^{-1}\left\{\sum_{i=1}^{n}\int_0^\tau(\beta-\beta_0)^T Z_i(\upsilon)\mathrm{d}N_i(\upsilon)\right\}-n^{-1}\sum_{i=1}^{n}\int_0^\tau\left\{\log\frac{S_n^{(0)}(\beta,\upsilon)}{S_n^{(0)}(\beta_0,\upsilon)}\right\}\mathrm{d}N_i(\upsilon)\}.$$

By Condition A2 and Lemma 1, $R(\beta)$ converges in probability to

$$\mathscr{R}(\beta)=(\beta-\beta_0)^T\int_0^\tau s^{(1)}(\beta_0,\upsilon)\mathrm{d}H_0(\upsilon)-\int_0^\tau\left\{\log\frac{s^{(0)}(\beta,\upsilon)}{s^{(0)}(\beta_0,\upsilon)}\right\}s^{(0)}(\beta_0,\upsilon)\mathrm{d}H_0(\upsilon).$$

By the facts that $s^{(1)}(\beta, x) = \partial s^{(0)}(\beta, x)/\partial\beta$ and $s^{(2)}(\beta, x) = \partial s^{(1)}(\beta, x)/\partial\beta$, one can prove that $\partial\mathcal{R}(\beta_0)/\partial\beta = 0$. With $\Omega(\beta_0) = -\partial^2\mathcal{R}(\beta_0)/\partial\beta\partial\beta^T$ being a positive-definite matrix by Condition A3, we see that $\mathcal{R}(\beta)$ is concave and maximized at $\beta = \beta_0$. The consistency of $\hat{\beta}_n$ is proved using a classic convex analysis theorem; see, e.g., Andersen & Gill (1982), Pons & de Turckheim (1988) or Lin et al. (2000).

Proof of Theorem 2

Let

$$n^{-1/2}U_{\mathrm{m}}(\beta_0)=n^{-1/2}\sum_{i=1}^{n}\tilde{U}_i(\beta_0)-n^{-1}\sum_{i=1}^{n}\int_0^\tau n^{1/2}\{\overline{Z}(\beta_0,\upsilon)-\overline{z}(\beta_0,\upsilon)\}\mathrm{d}M_i(\beta_0,\upsilon).$$

The second term converges in probability to zero, because $n^{1/2}\{\bar{Z}(\beta_0, \upsilon) - \bar{z}(\beta_0, \upsilon)\}$ converges weakly to a process with a continuous sample path and both $n^{-1}\Sigma_i N_i(x)$ and $n^{-1}\Sigma_i \Lambda_{\mathrm{m},i}(\beta_0, x)$ converge to $\phi(x)$, as shown in Lemma 1, under Conditions 1, A1, A2, A4 and A5. Thus, $n^{-1/2}U_{\mathrm{m}}(\beta_0)$ has the same limiting distribution as $n^{-1/2}\sum_{i=1}^{n}\tilde{U}_i(\beta_0)$.

A central limit theorem for $n^{-1/2}\sum_{i=1}^{n}\tilde{U}_i(\beta_0)$ can be proved by using central limit theorems from Doukhan et al. (1994) and Merlevede & Peligrad (2000) for strongly mixing random variables, upon assuming Condition A6 and that lim inf $n^{-1/2}E(\sum_{i=1}^{n}\tilde{U}_i)^{\otimes 2}>0$. Hence we can conclude that $n^{-1/2}\sum_{i=1}^{n}\tilde{U}_i(\beta_0)$ converges in distribution to a normal variable with mean zero and covariance matrix $\Phi(\beta_0)=\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}E\{\tilde{U}_i(\beta_0)\tilde{U}_j^{\mathrm{T}}(\beta_0)\}$.

Using the expansion $n^{-1/2}U_{\mathrm{m}}(\hat{\beta}_n) = n^{-1/2}U_{\mathrm{m}}(\beta_0) + n^{-1}U_{\mathrm{m}}(\beta^\dagger)n^{1/2}(\hat{\beta}_n - \beta_0) = 0$, where $\beta^\dagger$ is between $\hat{\beta}_n$ and β_0, we obtain $n^{1/2}(\hat{\beta}_n - \beta_0) = \{-n^{-1}U_{\mathrm{m}}(\beta^\dagger)\}^{-1}n^{-1/2}U_{\mathrm{m}}(\beta_0)$ where

$$n^{-1}U_{\mathrm{m}}(\beta)=\partial U_{\mathrm{m}}(\beta)/\partial\beta=n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\frac{S_n^{(2)}(\beta,\upsilon)}{S_n^{(0)}(\beta,\upsilon)}-\frac{S_n^{(1)}(\beta,\upsilon)^{\otimes 2}}{S_n^{(0)}(\beta,\upsilon)^2}\right\}\mathrm{d}N_i(\upsilon).$$

The stochastic equicontinuity of $n^{-1}U_{\mathrm{m}}(\beta)$ for $\beta \in \mathcal{B}^{-}$ holds by the uniform law of convergence in Lemma 2. This gives that for $\beta \in \mathcal{B}^{-}$, $n^{-1}U_{\mathrm{m}}(\beta) - \Omega(\beta_0)$ converges uniformly to 0 as $n \to \infty$ since $n^{-1}U_{\mathrm{m}}(\beta_0)$ converges to $\Omega(\beta_0)$. Thus, $n^{1/2}(\hat{\beta}_n - \beta_0)$ converges in distribution to a zero-mean normal variable with covariance $\Omega(\beta_0)^{-1}\,\Phi(\beta_0)\Omega\,(\beta_0)^{-1}$. The existence of $\Omega(\beta_0)^{-1}$ is guaranteed by Condition A3.

Proof of Theorem 3

The uniform convergence in x of $\hat{H}_0(\beta_0, x)$ in (4) to $H_0(x)$ follows from the uniform convergence of $S_n^{(0)}$ and $n^{-1}\,\Sigma_i N_i(x)$ in Lemmas 1 and 2. Since the derivative of $\hat{H}_0(\beta, x)$ with respect to β is uniformly bounded for large n and $\beta \in \mathcal{B}^{-}$, the consistency of $\hat{\beta}_n$ implies that $\hat{H}_0\,(\hat{\beta}_n, x)$ converges uniformly to $H_0(x)$. As for the weak convergence, through a simple decomposition we have

$$n^{1/2}\{\widehat{H}_0(\widehat{\beta}_n,x)-H_0(x)\}=n^{1/2}\{\widehat{H}_0(\widehat{\beta}_n,x)-\widehat{H}_0(\widehat{\beta}_0,x)\}+n^{-1/2}\sum_{i=1}^{n}\int_0^{x}S_n^{(0)}(\beta_0,\upsilon)^{-1}\mathrm{d}M_i(\beta_0,\upsilon).$$

By Taylor expansion, we can see that the first term on the right-hand side equals $(\partial/\partial\beta)\hat{H}_0(\beta^{\dagger}, x)n^{1/2}(\hat{\beta}_n - \beta_0)$ with $\beta^{\dagger}$ between $\hat{\beta}_n$ and β_0, where

$$(\partial/\partial\beta)\widehat{H}_0(\beta^{\dagger},x)=-n^{-1}\sum_{i=1}^{n}\int_0^{x}\frac{S_n^{(1)}(\beta^{\dagger},x)}{S_n^{(0)}(\beta^{\dagger},x)^2}\mathrm{d}N_i(x),$$

and further converges uniformly to a deterministic process $-\int_0^x \overline{z}(\beta_0,\upsilon)\mathrm{d}H_0(x)$. Since $\hat{\beta}_n$ is finite-dimensional and hence tight, the first term is asymptotically tight. The second term is approximated by $n^{-1/2}\sum_{i=1}^{n}\int_0^x s^{(0)}(\beta_0,\upsilon)^{-1}\mathrm{d}M_i(\beta_0,\upsilon)$; this converges weakly to a Gaussian process by Condition A6 and the convergence to finite-dimensional distributions, which can be proved by Theorem 1.3 in Merlevède & Peligrad (2000). Thus, $n^{1/2}\{\hat{H}_0(\hat{\beta}_n, x) - H_0(x)\}$ is uniformly asymptotically equivalent to $n^{-1/2}\sum_{i=1}^{n}\Xi_i(x)$. Weak convergence now follows from finite-dimensional limit theorems for strongly mixing processes and the tightness of the two terms.

References

Aalen OO, Fosen J, Weedon-FeKæjr H, Borgan Ø, Husebye E. Dynamic analysis of multivariate failure time data. Biometrics. 2004; 60:764–73. [PubMed: 15339300]

Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. Ann Statist. 1982; 10:1100–20.

Andrews DWK. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. Econometrica. 1991; 59:817–58.

Bagai I, Prakasa Rao BLS. Estimation of the survival function for stationary associated processes. Statist Prob Lett. 1991; 12:385–91.

Borgan Ø, Fiaccone RL, Henderson R, Barreto ML. Dynamic analysis of recurrent event data with missing observations, with application to infant diarrhoea in Brazil. Scand J Statist. 2007; 34:53–69.

Bradley RC. Basic properties of strong mixing conditions: A survey and some open questions. Prob Surveys. 2005; 2:107–44.

Cai Z. Kernel density and hazard rate estimation for censored dependent data. J Mult Anal. 1998; 67:23–34.

Cai Z. Estimating a distribution function for censored time series data. J Mult Anal. 2001; 78:299–318.

Cai Z, Roussas GG. Uniform strong estimation under α-mixing, with rates. Statist Prob Lett. 1992; 15:47–55.

Cai Z, Roussas GG. Kaplan-Meier estimator under association. J Mult Anal. 1998; 67:318–48.

Cleveland WS, Devlin SJ. Locally weighted regression: An approach to regression analysis by local fitting. J Am Statist Assoc. 1988; 83:596–610.

Cox, DR. The statistical analysis of dependencies in point processes. In: Lewis, PAW., editor. In Stochastic Point Processes. New York: John Wiley; 1972a. p. 55-66.

Cox DR. Regression models and life-tables (with Discussion). J R Statist Soc B. 1972b; 34:187–220.

Dabrowska DM, Ho WT. Estimation in a semiparametric modulated renewal process. Statist Sinica. 2006; 16:93–119.

Daley, DJ.; Vere-Jones, D. Introduction to the Theory of Point Processes Vol 1: Elementary Theory and Methods. 2nd. New York: Springer; 2003.

Doukhan P, Massart P, Rio E. The functional central limit theorem for strongly mixing processes. Ann Inst Henri Poincaré (B). 1994; 30:63–82.

Fosen J, Ferkingstad E, Borgan Ø, Aalen OO. Dynamic path analysis—A new approach to analyzing time-dependent covariates. Lifetime Data Anal. 2006; 12:143–67. [PubMed: 16817006]

Gandy A, Jensen U. A non-parametric approach to software reliability. Appl Stoch Mod Bus Indust. 2004; 20:3–15.

Hall P, Horowitz JL. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica. 1996; 64:891–916.

Hansen BE. Consistent covariance matrix estimation for dependent heterogeneous processes. Econometrica. 1992; 60:967–72.

Hougaard, P. Analysis of Multivariate Survival Data. New York: Springer; 2000.

Hu F, Kalbfleisch JD. The estimating function bootstrap (with Discussion). Can J Statist. 2000; 28:449–99.

Hu F, Zidek JV. A bootstrap based on the estimating equations of the linear model. Biometrika. 1995; 82:263–75.

Huang Y, Chen YQ. Marginal regression of gaps between recurrent events. Lifetime Data Anal. 2003; 9:293–303. [PubMed: 14649847]

Künsch HR. The jackknife and the bootstrap for general stationary observations. Ann Statist. 1989; 17:1217–41.

Lahiri, SN. Resampling Methods for Dependent Data. New York: Springer; 2003.

Lawless J, Nadeau C. Some simple robust methods for the analysis of recurrent events. Technometrics. 1995; 37:158–68.

Lele SR. Jackknifing linear estimating equations: Asymptotic theory and applications in stochastic processes. J R Statist Soc B. 1991a; 53:253–67.

Lele, SR. Resampling using estimating equations. In: Godambe, VP., editor. In Estimating Functions. Oxford Clarendon Press; 1991b. p. 295-304.

Lele SR. Impact of bootstrap on the estimating functions. Statist Sci. 2003; 18:185–90.

Leonenko NN, Sakhno LM. On the Kaplan–Meier estimator of long-range dependent sequences. Statist Infer Stoch Proces. 2001; 4:17–40.

Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. J R Statist Soc B. 2000; 62:711–30.

Lin DY, Wei LJ, Ying Z. Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika. 1993; 80:557–72.

Lin F, Fine JP. Pseudomartingale estimating equations for modulated renewal process models. J R Statist Soc B. 2009; 71:3–23.

Lin SC, Nicolelis M. Neuronal ensemble bursting in the basal forebrain encodes salience irrespective of valence. Neuron. 2008; 59:138–9. [PubMed: 18614035]

Luo X, Huang CY. Analysis of recurrent gap time data using the weighted risk-set method and the modified within-cluster resampling method. Statist Med. 2011; 30:301–11.

Merlevède F, Peligrad M. The functional central limit theorem under the strong mixing condition. Ann Prob. 2000; 28:1336–52.

Oakes D, Cui L. On semiparametric inference for modulated renewal processes. Biometrika. 1994; 81:83–90.

Peligrad M. On the blockwise bootstrap for empirical processes for stationary sequences. Ann Prob. 1998; 26:877–901.

Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. J Am Statist Assoc. 1993; 88:811–20.

Politis DN, Romano JP. A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. Ann Statist. 1992; 20:1985–2007.

Pons O, de Turckheim E. Cox's periodic regression model. Ann Statist. 1988; 16:678–93.

Shao Q, Yu H. Bootstrapping the sample means for stationary mixing sequences. Stoch Proces Appl. 1993; 48:175–90.

Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol. 2005; 93:1074–89. [PubMed: 15356183]

Yu H. A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences. Prob Theory Rel Fields. 1993; 95:357–70.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

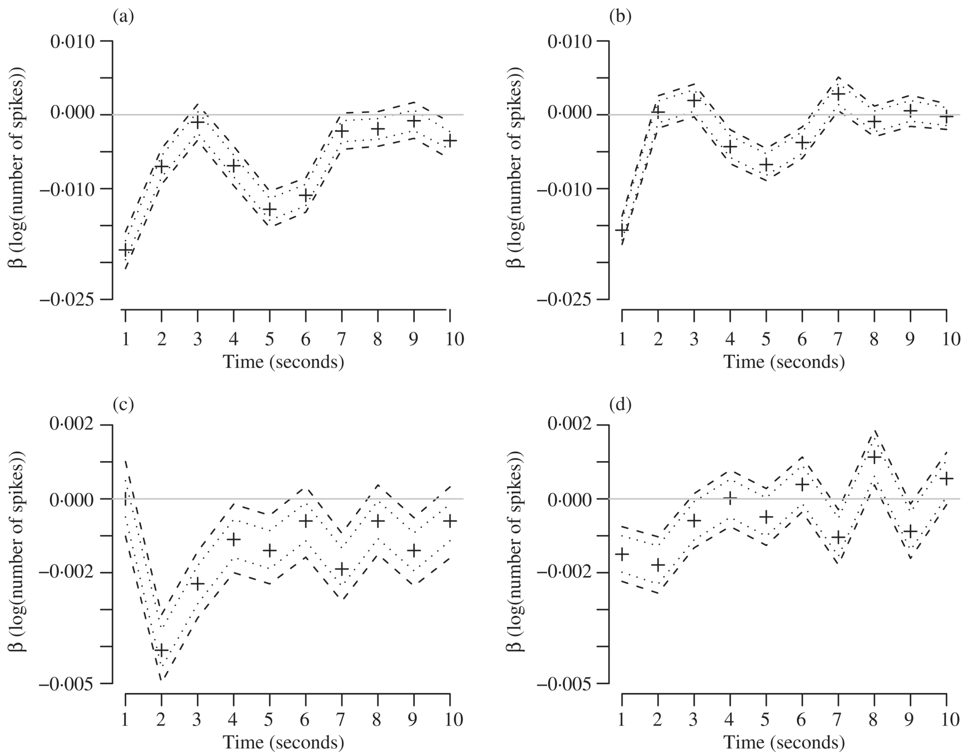

Fig. 1.

Point estimates (+) together with robust (dashed) and naive (dotted) 0·95 confidence intervals for the impacts of two peer cells, with or without the target neuron's spiking history included as covariates: (a) peer cell 1 without target neuron history; (b) peer cell 1 with target neuron history; (c) peer cell 2 without target neuron history; (d) peer cell 2 with target neuron history.

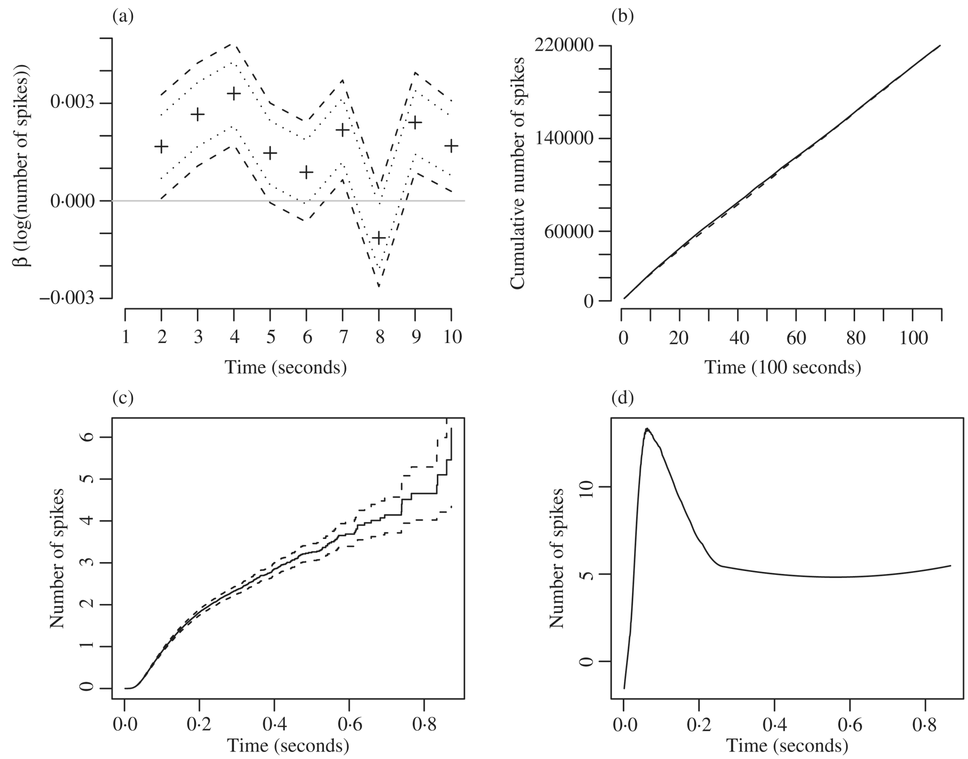

Fig. 2.
(a) Impact of target neuron's spiking history; (b) observed (solid) and estimated (dashed) cumulative number of spikes; (c) Breslow-type estimation of the cumulative baseline rate function (solid) together with 0·95 confidence interval (dashed); (d) naive estimation of the baseline rate function, assuming model (14).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Simulation results for model (10) when $\gamma_r = 0$ for all *r* in model (9); entries have been multiplied by 10^2

d	α_0	n	Bias	Σ_0	$\hat{\Sigma}_0$ (5)	$\widehat{\sum}_{\mathrm{c}}^{d}$ (8)	$\widehat{\sum}_{\mathrm{b}}^{N}$ (6)	$\widehat{\sum}_{\mathrm{b}}^{M}$ (7)	π_0 (5)	π_{c}^{d} (8)	π_{b}^{N} (6)	π_{b}^{M} (7)
0	–	200	1-07	8-72	8-20	7-90	7-97	7-69	93-0	92-4	92-4	91-9
		400	1-18	4-32	4-04	3-97	3-96	3-89	94-3	93-9	93-7	93-7
		1000	0-86	1-49	1-60	1-59	1-59	1-58	95-9	95-6	95-6	95-8
		2000	0-69	0-83	0-80	0-79	0-79	0-79	94-7	94-4	94-1	94-2
2	0-75	200	2-60	8-76	7-38	7-78	8-24	7-40	93-7	91-9	94-2	93-5
		400	1-57	4-28	3-63	4-07	4-24	3-78	93-2	92-2	94-7	92-9
		1000	1-07	1-58	1-44	1-68	1-77	1-56	94-2	95-5	96-4	95-4
		2000	1-15	0-83	0-72	0-84	0-91	0-79	93-0	94-6	95-5	93-8
2	0-5	200	6-81	10-70	7-44	9-27	9-61	7-98	89-9	91-0	92-5	90-1
		400	4-54	4-77	3-66	4-88	4-96	4-15	91-0	93-0	94-9	92-3
		1000	1-13	2-31	1-44	2-02	2-10	1-75	87-7	92-9	93-5	91-4
		2000	0-60	1-09	0-72	1-03	1-10	0-92	89-1	94-2	94-7	92-1
5	0-75	200	3-04	10-18	7-39	8-00	8-49	7-43	90-7	86-0	91-5	90-2
		400	2-09	5-22	3-63	4-40	4-37	3-81	89-4	90-0	92-6	90-2
		1000	0-97	2-17	1-44	1-94	1-89	1-61	88-9	92-4	93-2	89-8
		2000	0-59	1-01	0-72	0-98	1-00	0-83	90-1	94-0	94-5	92-0
5	0-5	200	9-08	14-55	7-51	10-94	10-20	8-34	83-3	83-5	88-8	84-1
		400	6-49	6-63	3-66	6-05	5-24	4-31	85-9	91-1	90-7	88-7
		1000	2-99	2-91	1-44	2-66	2-37	1-91	82-1	92-4	91-9	87-9
		2000	1-01	1-37	0-72	1-38	1-29	1-02	85-0	94-4	94-3	90-3

Bias is defined by $\hat{\beta_n} - \beta_0$; Σ_0, empirical variance; π, coverage probability.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Simulation results for model (10) when $\gamma_r = 10^{-r}$ and $R = 10$ in model (9); entries have been multiplied by 10^2

Size	a_0	n	Bias	Σ_0	$\hat{\Sigma_0}$ (5)	$\widehat{\sum}_c^d$ (8)	$\widehat{\sum}_b^N$ (6)	$\widehat{\sum}_b^M$ (7)	π_0 (5)	π_c^d (8)	π_b^N (6)	π_b^M (7)
5	0·75	400	−6·84	4·36	3·56	3·60	3·96	3·50	92·0	91·1	93·4	92·1
		1000	−3·06	1·56	1·37	1·44	1·54	1·39	91·6	92·9	93·4	92·0
		2000	−1·63	0·75	0·68	0·72	0·77	0·70	93·7	94·2	94·5	93·8
		4000	−1·13	0·36	0·34	0·36	0·39	0·35	93·7	94·4	94·9	94·7
5	0·5	400	−7·77	4·32	3·38	3·45	3·76	3·35	90·1	89·9	91·4	89·2
		1000	−3·61	1·51	1·30	1·36	1·46	1·32	91·8	92·2	93·7	91·6
		2000	−1·95	0·75	0·64	0·68	0·73	0·67	92·9	93·8	94·7	93·6
		4000	−1·29	0·36	0·32	0·35	0·37	0·34	93·3	94·8	95·3	94·2
10	0·75	400	−13·20	4·42	3·59	3·79	4·03	3·58	87·9	85·9	82·5	86·9
		1000	−5·12	1·94	1·37	1·55	1·59	1·42	87·4	89·2	90·2	87·6
		2000	−2·58	0·81	0·68	0·78	0·81	0·72	91·7	93·0	93·4	91·7
		4000	−0·83	0·39	0·34	0·40	0·42	0·36	93·0	95·2	95·4	94·0
10	0·5	400	−15·70	4·85	3·40	4·00	3·91	3·50	81·9	83·6	81·7	81·7
		1000	−6·82	1·91	1·30	1·52	1·52	1·36	85·3	87·7	87·9	85·8
		2000	−3·38	0·81	0·64	0·76	0·77	0·68	90·3	92·5	92·9	91·5
		4000	−1·30	0·39	0·32	0·38	0·39	0·35	91·9	94·8	94·8	93·5

Bias is defined by $\hat{\beta_n} - \beta_0$; Σ_0, empirical variance; π, coverage probability.